Peabody, Daniel (EGLE)

From: Saric, James <saric.james@epa.gov>
Sent: Wednesday, April 8, 2020 10:10 AM
To: Peabody, Daniel (EGLE); Roberts, Keegan; Ruhala, Sydney (EGLE); john kern; Keiser, Jeff/MKE; Von Wallmenich, Theo/DET; Thompson, Robert A.; Johnson, Shannon D; Draper, Cynthia E
Cc: Roth, Charles; Canar, John; Wood, Nicole; Frey, Rebecca
Subject: Statistical Analysis of Split Samples
Attachments: FIELDS Analysis of Replicate Lab Samples_Final.docx

Attached is the final report developed by EPA's FIELDS group summarizing the statistical analysis of the split samples taken from the Kalamazoo River OU5, Area 4 Trowbridge Impoundment.

Thanks

Jim Saric
Remedial Project Manager
US EPA Region 5, Chicago
(312) 886 - 0992

FIELDS Analysis of Replicate Lab Samples for [tPCB]

John Canar, FIELDS Group
Chuck Roth, FIELDS Group
RPM: Jim Saric

4/8/20

Introduction

The results measuring total [PCB] from RI sampling (historic samples from 1993-2007) and SRI sampling (new samples from 2015 on) appear to represent different population distributions with historic samples showing higher [PCB] than the newer data. Several comparisons of historic to newer data using nearby samples as pairs demonstrate the change in distribution. Additionally, a more recent sampling where splits were taken appear to suggest the newer sample analysis methods (USEPA approved QAPPs) are measuring total [PCB] differently. However, both observations have certain weaknesses and the specific cause(s) has not been isolated. Comparison of neighbors likely introduces variability based on the distances between sample location, while the recent split samples used different QC assumptions and sample concentrations were generally low. Although there are no experimental methods to compare historic and newer data sets, the quality of sampling and quantification methods can be compared to assure the methods being used in the future are representing as close to the true [PCB] as possible.

In October 2019 samples were collected from the Area 4 OU5 floodplain and 30 samples, collected from 10 locations, were split and sent to two labs for ARACLOR analysis under an EPA approved QAPP: Pace Labs (Wood) and ASB (EPA lab). When it was determined there were likely significant differences between the two labs, samples were then also sent to an additional lab, CLP (EPA contract lab) under a different although EPA-approved QAPP. EGLE also conducted PCB analysis of the split samples using Vista laboratory for congener analysis, under a different non-EPA approved QAPP. For clarification, the CLP lab analyzed for AROCLORS and the EGLE lab analyzed for PCB congeners. PCBs were then summed for a measure of total PCBs (tPCB) for each lab. The purpose of the analyses was to determine if there are differences between the labs and if there is a pattern to those differences.

Statistically the results were used to test the hypothesis that there is no difference between the labs. If a difference is found then the labs were compared to each other to determine which labs were different and which labs were the same as each other. The implications of the difference would warrant further investigation. Since all of the samples were prepared the same way in the field the methods of extraction, testing, and interpretation of the results need to be compared to determine why the results differ from each other.

Methods

Repeated Measures ANOVA. A Repeated Measures ANOVA was used to test for differences between lab results for each sample. Since the heterogeneity within each sample jar is expected to be low there should not be large differences between labs for each sample. Each lab's measure of a sample represents a repeated measure of the sample. The null hypothesis is that there are no differences between the measurement of each sample by each lab (result of lab A for Sample 1 = lab B = Lab C, result of lab A for Sample 2 = lab B = Lab C …). The alternate hypothesis is that one or more of the labs are significantly different. A series of post hoc tests can then be used to determine which lab or labs are different. The Repeated Measures test can be performed parametrically if a test of normality finds the sampling distributions to be likely to be normal. If not, a Repeated Measure test on the ranks can be used. The same applies to the post-hoc tests. For non-parametric pairwise comparisons the Tukey-Kramer will be used. The Tukey-Kramer is a pairwise comparisons method that controls the maximum experimentwise error rate for both equal and unequal sample sizes.

Post-hoc Tests. If significant differences between the labs are found, a series of post hoc tests are performed to identify the labs that are different. Since each lab sample is from the same sample jar the results are dependent, meaning that, if two labs are getting the same results, the mathematical difference for each sample should equal (or at least near) zero. With the Sign test, the null hypothesis is tested by examining the frequency that lab A is greater than lab B. The null hypothesis is that the frequency is equal, Lab A is higher 50% of the time and lab B is higher 50% of the time. The greater the deviation from 50/50 the more likely the labs are getting different results. The Sign test can be used on the raw results or on the ranks of the results. A paired t-test can also be used to test for differences between lab results. The null hypothesis is that the mean (or median for nonparametric) of the difference for each sample equals zero.

Data: The results from this sampling have results ranging from 0.002 ppm to 62.23 ppm. A considerable number of the results were either non-detects or very low results (60 out of 119 are less than 1.0). The differences between labs in the range below 1 ppm are driven as much by difference in limits of detection as they are by differences between labs. In order to examine the differences between labs at a data range where the differences are important, a separate dataset was generated with samples removed where all labs' results were less than 1ppm (<1 ppm was selected based on the difference in relative importance to decision-making, but 65% of the samples eliminated were either non-detects or results < 0.07 ppm, with 0.652ppm highest result removed). Both datasets were statistically tested for differences. For comparison, the results from the entire dataset can be found in the appendix, slides 4-19. However, we consider the results without the bias of the non-detects to be more useful in explaining the differences between lab results. The differences at very low concentrations may more reflect differences due to differing limits of detection and random variation. A list of the results and those samples removed can also be found in the appendix, slides 1 & 2.

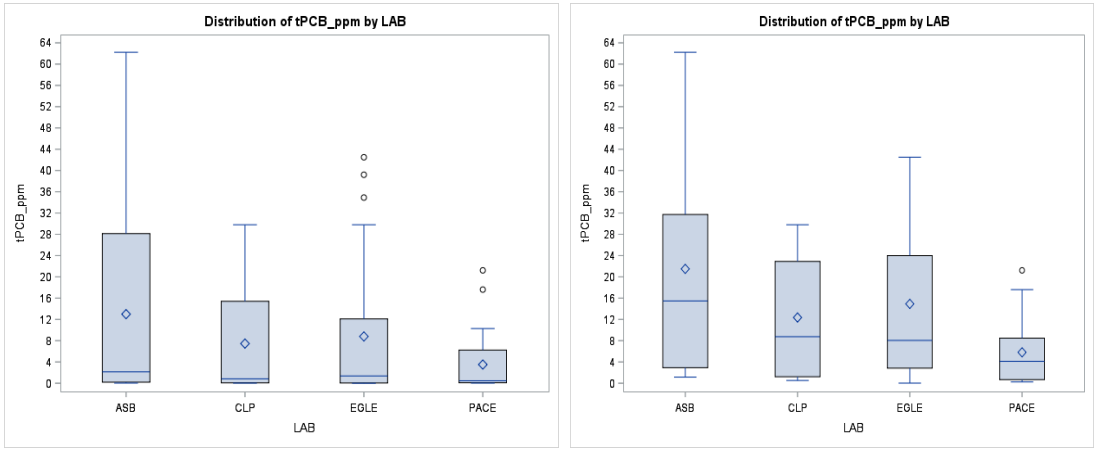
Individual AROCLORS: Each AROLCLOR with significant results was also examined using the above protocol. 1248 and 1254 make up the bulk of the tPCB, with 1260 having low concentrations (non-detect to 2 ppm).

Results

Descriptive Statistics: The descriptive statistics show the differences between each lab. ASB generally has the highest results and the Pace Lab is generally the lowest (see Figures 1a and 1b). The pattern is a bit more obvious where all labs' results for a sample are > 1 ppm.

The Q-Q plots of the distributions for each lab show that the ranking for each lab is consistent at lower and higher concentrations, ASB is consistently higher and Pace is consistently lower (see Figure 2). Additionally, the differences between labs appear to be non-linear, with differences at higher concentration greater than at lower concentrations.

Statistical tests were performed using all results and on a subset of the data where samples were removed if all results were less than 1 ppm. Figure 3 shows the descriptive statistics for the subset data.



Figures 1a and 1b. Box plots of the results for each lab: 1a-All results, 1b-Samples where all lab's results are > 1ppm.

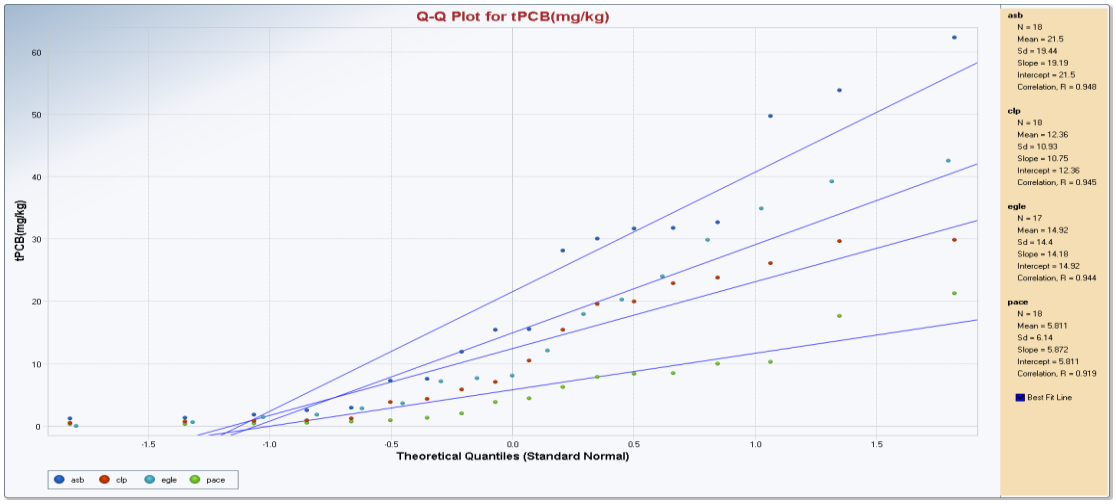


Figure 2. Q-Q plots of all results for each lab.

Descriptive statistics for total PCB by Lab
PCB split sampling project (only highest 18 quadruplicates)
Kalamazoo River Area 4 Floodplain samples

The MEANS Procedure

Analysis Variable : tPCB_ppm					
LAB	N Obs	Mean	Median	Minimum	Maximum
ASB	18	21.4996667	15.4715000	1.1470000	62.2300000
CLP	18	12.3588889	8.7500000	0.5200000	29.8000000
EGLE	18	14.9188294	8.0500000	0.0211000	42.5000000
PACE	18	5.8107222	4.1150000	0.2440000	21.2200000

Figure 3. Descriptive statistics for each lab with all samples where all results are < 1ppm

Repeated Measures ANOVA Results: Both datasets were tested for normality with goodness of fit tests (results are located in the appendix, slides 8 & 22) and were found to not be normally distributed. Following are the results of the nonparametric results (see Figure 4). (For the parametric results refer to the appendix, slides 11-14 & 25-28.) The results for both datasets are highly significant indicating that one or more of the lab results are significantly different.

Proc GLM and post-hoc tests of Ranked tPCB Lab differences
repeated measures ANOVA tested as a two-way ANOVA
PCB split sampling project (only highest 18 quadruplicates)
Kalamazoo River Area 4 Floodplain samples

The GLM Procedure

Dependent Variable: tPCB_ppm_r Rank for Variable tPCB_ppm

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	27857.62745	1392.88137	35.52	<.0001
Error	50	1960.87255	39.21745		
Corrected Total	70	29818.50000			

R-Square	Coeff Var	Root MSE	tPCB_ppm_r Mean
0.934240	17.39551	6.262384	36.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sample_Name	17	24082.25000	1416.60294	36.12	<.0001
LAB	3	3775.37745	1258.45915	32.09	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sample_Name	17	23989.94363	1411.17315	35.98	<.0001
LAB	3	3775.37745	1258.45915	32.09	<.0001

Figure 4. Results of the Ranked Repeated Measures ANOVA for samples with all results > 1ppm.

Post-hoc Results: The Tukey-Kramer test for comparing means was used to separate which Labs were significantly different (Figure 5). The test incorporates experimentwise error to adjust for multiple comparisons (see Appendix, slide 3, for an explanation of experimentwise error). The post-hoc tests determine that ASB lab is significantly higher than all other labs. Pace lab was found to be significantly lower than all other labs. The EGLE congener results and the CLP lab were not significantly different. A separate post-hoc test, the Sign Test found the same results (Figure 6).

Plotting the residuals for each paired comparison is provided to examine how the labs differ from each other (Figures 7a-f). Generally, if the two samples represent the same population, the distribution of the residuals is expected to be random (no pattern). In the case of the sign test that would be defined as equal numbers of (+) differences and (-) differences. The comparison of the CLP AROCLOR to the EGLE

congener data sets, which were found to not be significantly different, shows no pattern and relatively similar positive and negative differences (Figure 7a). All of the other comparison show either an upward trend as concentrations get higher, a predominance of positive sign (or negative if the subtraction were opposite), or both (Figures 7b-f). The graphics support the conclusion that ASB is significantly higher than all other labs, Pace is significantly lower than all labs, and that CLP and EGLE congeners are not significantly different.

Proc GLM and post-hoc tests of Ranked tPCB Lab differences
repeated measures ANOVA tested as a two-way ANOVA
PCB split sampling project (only highest 18 quadruplicates)
Kalamazoo River Area 4 Floodplain samples

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

LAB	tPCB_ppm_r LSMEAN	LSMEAN Number
ASB	45.3055556	1
CLP	35.3055556	2
EGLE	37.2532680	3
PACE	24.9722222	4

Least Squares Means for effect LAB
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: tPCB_ppm_r

i/j	1	2	3	4
1		<.0001	0.0023	<.0001
2	<.0001		0.7968	<.0001
3	0.0023	0.7968		<.0001
4	<.0001	<.0001	<.0001	

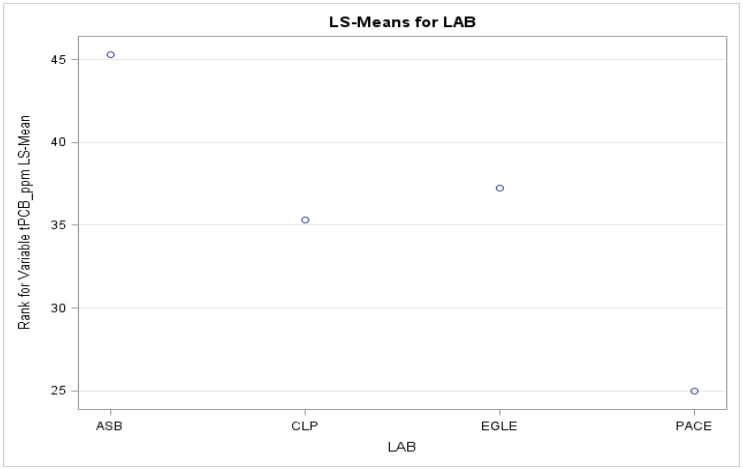


Figure 5. Results of the Tukey-Kramer post-hoc tests for the dataset with samples with all results <1 ppm removed.

p values	EGLE	ASB	CLP	Pace
EGLE	x			
ASB	**0.00636**	x		
CLP	0.3145	**3.8-06**	X	
Pace	**0.00636**	**3.8E-6**	**7.24E-05**	X

Figure 6. Results of the Sign Test for samples all >1ppm

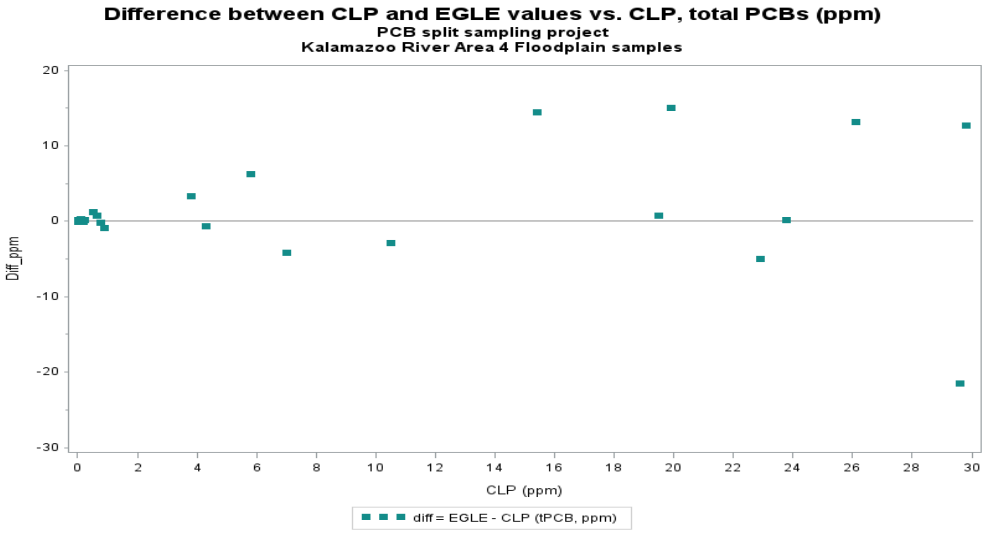


Figure 7a. Residuals for EGLE-CLP. No pattern and differences both positive and negative.

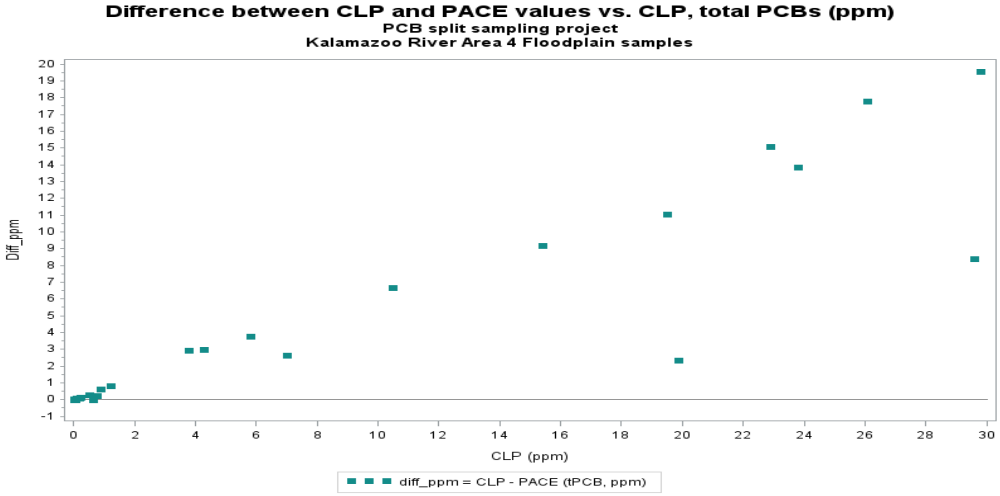


Figure 7b. Residuals for CLP-Pace. Pace always lower and greater differences at higher concentrations.

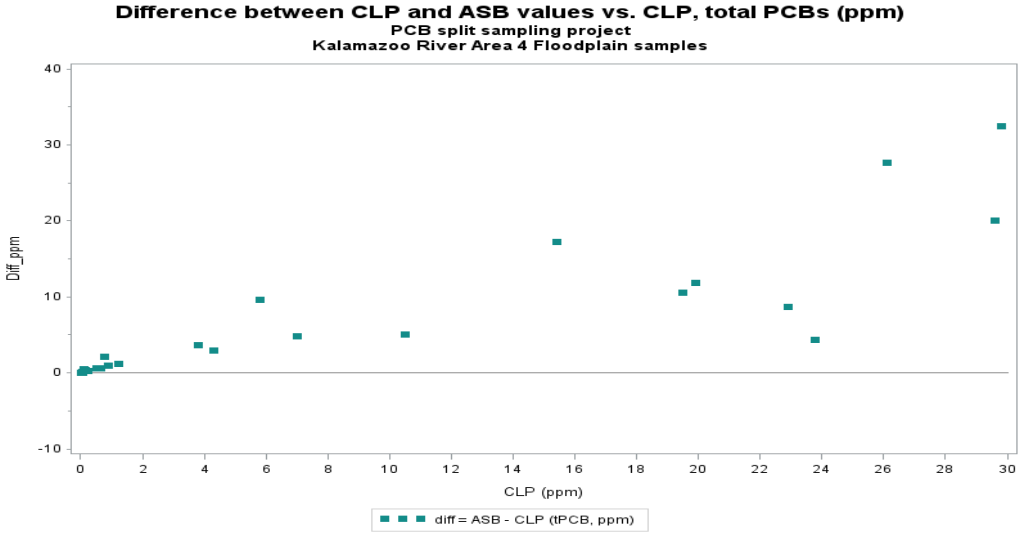


Figure 8c. Residuals for ASB-CLP. ASB always higher and greater differences at higher concentrations.

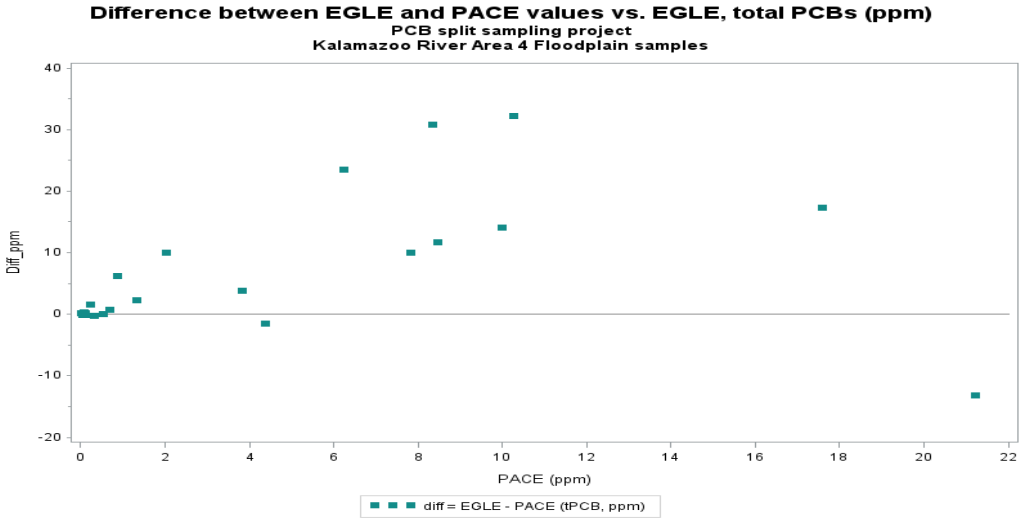


Figure 7d. Residuals for EGLE-Pace. Pace always lower, less of an upward trend at higher concentrations.

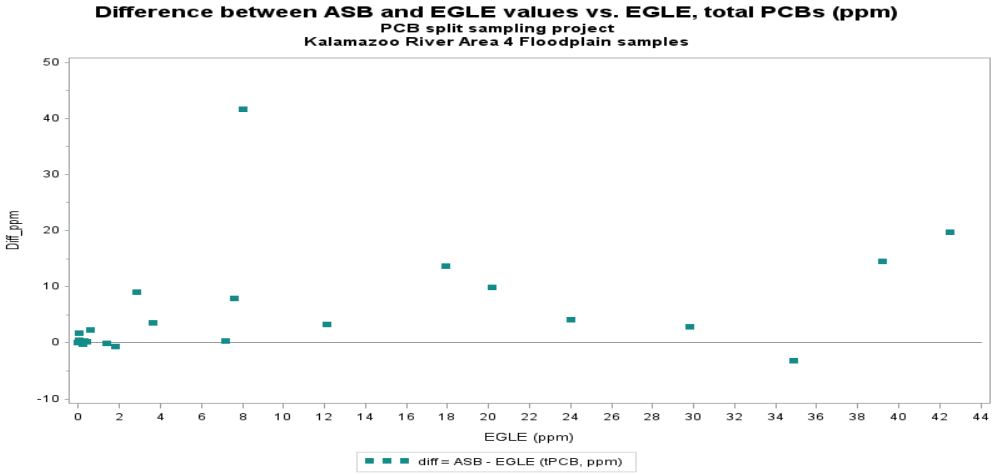


Figure7e. Residuals for ASB-EGLE. ASB nearly always higher but no pattern at higher concentrations.

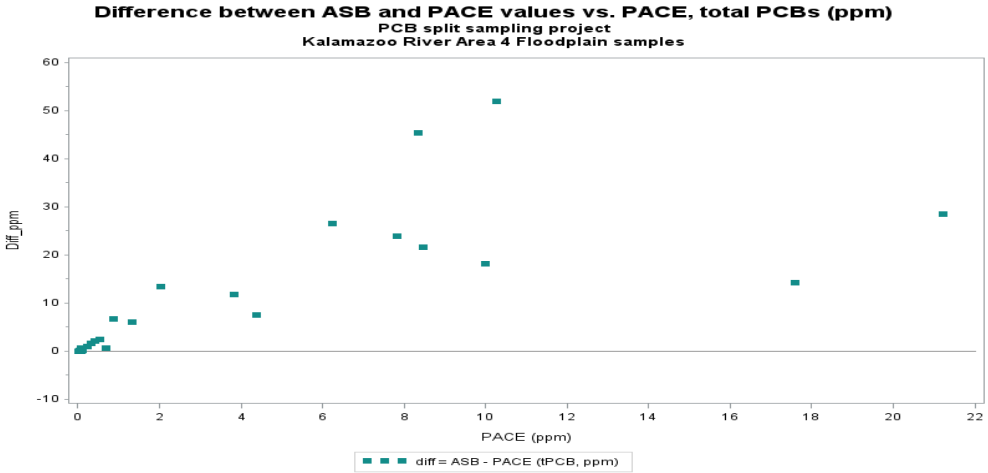


Figure 7f. Residuals for ASB-Pace. ASB always higher and a slight upward trend at higher concentrations.

Summary of Findings

Repeated Measures ANOVA found significant difference between labs

Means: ASB>EGLE>CLP>Pace

Tukey-Kramer post-hoc tests and Sign test

ASB>EGLE≈CLP>Pace

tPCB Comparisons: Thirty samples from the Area 4 floodplain were split and sent to the four labs referenced in this report. The split samples are expected to have a minimum of heterogeneity providing an opportunity for direct comparisons between lab and the coinciding preparation, analysis, and quantification methods. The implication of significant differences in lab results is that there are significant differences in lab preparation, analysis, and/or quantification methods. ASB lab was found to be significantly higher than all of the other labs, with the differences greater with higher concentrations. Pace lab was found to be significantly lower than all of the other labs, with increasing difference at higher concentrations. It is important to note that agreement does equal the "correct" answer. The actual concentration is not going to be a known. However, there needs to be better agreement between different labs, which will require determining the differences in lab extraction and quantification methods and adjusting the accepted methods to narrow the variability between the lab results.

Individual AROCLORS: Each AROCLOR with significant results was also examined (the EGLE results are congeners only). Both 1248 and 1254 follow the same pattern as the tPCB concentrations, with the same significant differences found with the Repeated Measures ANOVA (see appendix slides 35-94 for results). It is unlikely that differences are due to lab differences in quantifying the individual AROCLORS.

Table of the Results

Location	Sample Name	PCB	EGLE	ASB	CLP	PACE
A4-FPS-206	A4-FPS-206-24-33-0919	tPCB(mg/kg)	0.00023	0.071	0.033	0.051
A4-FPS-202	A4-FPS-202-24-34-0919	tPCB(mg/kg)	0.00211	0.0719	0.034	0.053
A4-FPS-206	A4-FPS-206-12-24-0919	tPCB(mg/kg)	0.00351	0.073	0.034	0.052
A4-FPS-203	A4-FPS-203-24-28-0919	tPCB(mg/kg)	0.0211	1.817	0.91	0.32
A4-FPS-204	A4-FPS-204-24-31-0919	tPCB(mg/kg)	0.0274	0.0255	0.022	0.056
A4-FPS-202	A4-FPS-202-12-24-0919	tPCB(mg/kg)	0.0277	0.0306	0.038	0.087
A4-FPS-205	A4-FPS-205-24-33-0919	tPCB(mg/kg)	0.0327	0.0383	0.06	0.07
A4-FPS-211	A4-FPS-211-24-33-0919	tPCB(mg/kg)	0.0513	0.489	0.18	0.098
A4-FPS-207	A4-FPS-207-24-30-0919	tPCB(mg/kg)	0.118	0.2058	0.075	0.11
A4-FPS-206	A4-FPS-206-0-12-0919	tPCB(mg/kg)	0.223	0.0273	0.023	0.029
A4-FPS-209	A4-FPS-209-24-33-0919	tPCB(mg/kg)	0.253	0.5676	0.15	0.066
A4-FPS-210	A4-FPS-210-24-36-0919	tPCB(mg/kg)	0.314	0.504	0.209	0.099
A4-FPS-208	A4-FPS-208-24-31-0919	tPCB(mg/kg)	0.4	0.652	0.12	0.078
A4-FPS-207	A4-FPS-207-12-24-0919	tPCB(mg/kg)	0.559	2.923	0.75	0.52
A4-FPS-204	A4-FPS-204-12-24-0919	tPCB(mg/kg)	1.37	1.281	0.66	0.677

A4-FPS-209	A4-FPS-209-12-24-0919	tPCB(mg/kg)	1.79	1.147	0.52	0.244
A4-FPS-202	A4-FPS-202-0-12-0919	tPCB(mg/kg)	2.83	11.856	7	4.39
A4-FPS-205	A4-FPS-205-12-24-0919	tPCB(mg/kg)	3.63	7.243	4.3	1.318
A4-FPS-211	A4-FPS-211-12-24-0919	tPCB(mg/kg)	7.16	7.505	3.8	0.884
A4-FPS-203	A4-FPS-203-12-24-0919	tPCB(mg/kg)	7.61	15.543	10.5	3.84
A4-FPS-211	A4-FPS-211-0-12-0919	tPCB(mg/kg)	8.05	49.63	29.6	21.22
A4-FPS-208	A4-FPS-208-12-24-0919	tPCB(mg/kg)	12.1	15.4	5.8	2.03
A4-FPS-205	A4-FPS-205-0-12-0919	tPCB(mg/kg)	17.9	31.62	22.9	7.82
A4-FPS-209	A4-FPS-209-0-12-0919	tPCB(mg/kg)	20.2	30.06	19.5	8.47
A4-FPS-203	A4-FPS-203-0-12-0919	tPCB(mg/kg)	24	28.13	23.8	9.98
A4-FPS-210	A4-FPS-210-0-12-0919	tPCB(mg/kg)	29.8	32.66	15.4	6.24
A4-FPS-204	A4-FPS-204-0-12-0919	tPCB(mg/kg)	34.9	31.72	19.9	17.6
A4-FPS-207	A4-FPS-207-0-12-0919	tPCB(mg/kg)	39.2	53.76	26.1	8.35
A4-FPS-208	A4-FPS-208-0-12-0919	tPCB(mg/kg)	42.5	62.23	29.8	10.28
A4-FPS-210	A4-FPS-210-12-24-0919	tPCB(mg/kg)	n/d	2.469	1.22	0.41

Appendix (see attached PowerPoint. ForReportSplitComparisons2.pptx)

The appendix contains all of the analyses, figures, and tables performed by the FIELDS team regarding this data, including those in this report. Please contact Chuck Roth, Roth.Charles@epa.gov or John Canar, Canar.John@epa.gov with any questions regarding the analyses.

References

SAS v9.4 statistical software, Cary, NC: SAS Institute Inc.

Cody, R.P. and Smith, J.K., Applied Statistics and the SAS Programming Language, fourth edition. Prentice Hall, Upper Saddle River, New Jersey. 1997

Sokal, R.R. and Rohlf, F.J., Biometry, third edition. W. H. Freeman and Company, New York, New York. 1995